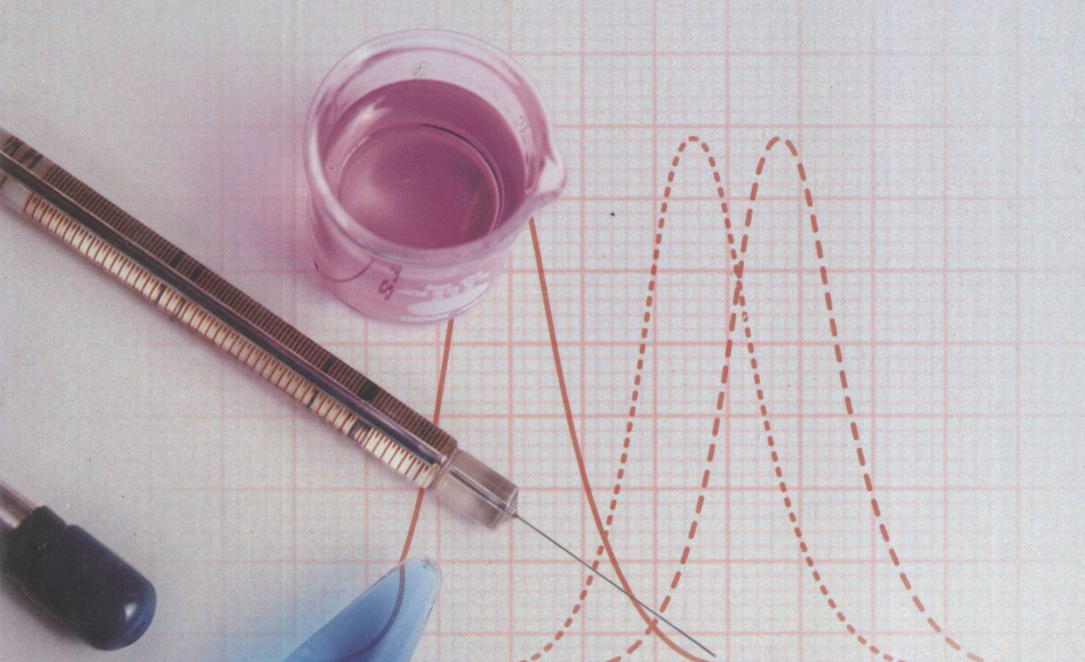
SECOND EDITION

Applied Mixed Models in Medicine



HELEN BROWN
ROBIN PRESCOTT

WILEY

STATISTICS IN PRACTICE

Applied Mixed Models in Medicine

Second Edition

Helen Brown

NHS Scotland
Edinburgh
UK

and

Robin Prescott

Medical Statistics Unit
Public Health Sciences
University of Edinburgh
UK

John Wiley & Sons, Ltd

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 42 McDougall Street, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging-in-Publication Data:

Brown, Helen, 1962–
Applied mixed models in medicine / Helen Brown and Robin Prescott. – 2nd ed.
p. ; cm.
Includes bibliographical references and index.
ISBN-13: 978-0-470-02356-3 (cloth : alk. paper)
ISBN-10: 0-470-02356-2 (cloth : alk. paper)
1. Medicine – Research – Statistical methods. 2. Medicine – Statistical methods. 3. Statistics. I. Prescott, Robin. II. Title.
[DNLM: 1. Medicine. 2. Statistics – methods. 3. Models, Statistical. WA 950 B878a 2006]
R853.S7B76 2006
610.72′7 – dc22

2005036661

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN-13: 978-0-470-02356-3 (HB)
ISBN-10: 0-470-02356-2 (HB)

Typeset in 10/12pt Photina by Laserwords Private Limited, Chennai, India
Printed and bound in Great Britain by TJ International, Padstow, Cornwall
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

Applied Mixed Models in Medicine

Second Edition

Statistics in Practice

Advisory Editor

Stephen Senn
University of Glasgow, UK

Founding Editor

Vic Barnett
Nottingham Trent University, UK

Statistics in Practice is an important international series of texts which provide detailed coverage of statistical concepts, methods and worked case studies in specific fields of investigation and study.

With sound motivation and many worked practical examples, the books show in down-to-earth terms how to select and use an appropriate range of statistical techniques in a particular practical field within each title's special topic area.

The books provide statistical support for professionals and research workers across a range of employment fields and research environments. Subject areas covered include medicine and pharmaceutics; industry, finance and commerce; public services; the earth and environmental sciences, and so on.

The books also provide support to students studying statistical courses applied to the above areas. The demand for graduates to be equipped for the work environment has led to such courses becoming increasingly prevalent at universities and colleges.

It is our aim to present judiciously chosen and well-written workbooks to meet everyday practical needs. Feedback of views from readers will be most valuable to monitor the success of this aim.

A complete list of titles in this series appears at the end of the volume.

Preface to Second Edition

Analysis of variance and regression analysis have for many years been the mainstay of statistical modelling. These techniques usually have as a basic assumption that the residual or error terms are independently and identically distributed. Mixed models are an important new approach to modelling which allows us to relax the independence assumption and take into account more complicated data structures in a flexible way. Sometimes this interdependence of observations is modelled directly in a mixed model. For example, if a number of repeated measurements are made on a patient, mixed models allow us to specify a pattern for the correlation between these measurements. In other contexts, such as the cross-over clinical trial, specifying that patient effects are normally distributed, rather than fixed as in the classical approach, induces observations on the same patient to be correlated.

There are many benefits to be gained from using mixed models. In some situations the benefit will be an increase in the precision of our estimates. In others, we will be able to make wider inferences. We will sometimes be able to use a more appropriate model which will give us greater insight into what underpins the structure of the data. However, it is only the relatively recent availability of software in versatile packages such as SAS® which has made these techniques widely accessible. It is now important that suitable information on their use becomes available so that they may be applied confidently on a routine basis.

Our intention in this book is to put all types of mixed model into a general framework and to consider the practical implications of their use. We aim to do this at a level that can be understood by applied statisticians and numerate scientists. Greatest emphasis is placed on skills required for the application of mixed models and interpretation of the results. An in-depth understanding of the mathematical theory underlying mixed models is not essential to gain these, but an awareness of the practical consequences of fitting different types of mixed models is necessary. While many publications are available on various aspects of mixed models, these generally relate to specific types of model and often differ in their use of terminology. Such publications are not always readily comprehensible to the applied statisticians who will be the most frequent users of the methods. An objective of this book is to help overcome this deficit.

Examples given will primarily relate to the medical field. However, the general concepts of mixed models apply equally to many other areas of application, e.g. social sciences, agriculture, official statistics. (In the social sciences mixed models are often referred to as 'multi-level' models.) Data are becoming easier to collect with the consequence that datasets are now often large and complex. We believe that mixed models provide useful tools for modelling the complex structures that occur in such data.

The second edition of this book retains the same structure as the first edition, but there are substantial changes to reflect the advances in software. All of the examples from the first edition have been reworked in Version 9 of SAS using, where relevant, the newly available `PROC GLIMMIX` for fitting generalised linear mixed models and categorical mixed models. There are also new sections on bioequivalence studies and cluster randomised trials, and a complicated laboratory experiment has been added to the examples. There has been a general updating throughout to reflect recent literature and software changes.

Chapter 1 provides an introduction to the capabilities of mixed models, defines general concepts and gives their basic statistical properties. Chapter 2 defines models and fitting methods for normally distributed data. Chapter 3 first introduces generalised linear models which can be used for the analysis of data which are binomial or Poisson or from any other member of the exponential family of distributions. These methods are then extended to incorporate mixed models concepts under the heading of generalised linear mixed models. The fourth chapter then examines how mixed models can be applied when the variable to be analysed is categorical. The main emphasis in these chapters, and indeed in the whole book, is on classical statistical approaches to inference, based on significance tests and confidence intervals. However, the Bayesian approach is also introduced, since it has several potential advantages and its use is becoming more widespread. Although the overall emphasis of the book is on the application of mixed models techniques, these chapters can also be used as a reference guide to the underlying theory of mixed models.

Chapters 5–7 consider the practical implications of using mixed models for particular designs. Each design illustrates a different feature of mixed models.

Multi-centre trials and meta-analyses are considered in Chapter 5. These are examples of hierarchical data structures and the use of a mixed model allows for any additional variation in treatment effects occurring between centres (or trials) and hence makes results more generalisable. The methods shown can be applied equally to any type of hierarchical data.

In Chapter 6 the uses of covariance pattern models and random coefficients models are described using the repeated measures design. These approaches take into account the correlated nature of the repeated observations and give more appropriate treatment effect estimates and standard errors. The material in this chapter will apply equally to any situation where repeated observations are made on the same units.

Chapter 7 considers cross-over designs where each patient may receive several treatments. In this design more accurate treatment estimates are often achieved by fitting patient effects as random. This improvement in efficiency can occur for any dataset where a fixed effect is 'crossed' with a random effect.

In Chapter 8 a variety of other designs and data structures are considered. These either incorporate several of the design aspects covered in Chapters 5–7 or have structures that have arisen in a more unplanned fashion. They help to illustrate the broad scope of application of mixed models. This chapter includes the new section on bioequivalence studies and cluster randomised trials.

Chapter 9 gives information on software available for fitting mixed models. Most of the analyses in the book are carried out using `PROC MIXED` in SAS, supplemented by `PROC GENMOD` and `PROC GLIMMIX`, and this chapter introduces the basic syntax for these procedures. This information should be sufficient for fitting most of the analyses described but the full SAS documentation should be referenced for those who wish to use more complex features. The SAS code used for most of the examples is supplied within the text. Additionally, the example datasets and SAS code may be obtained electronically from www.chs.med.ac.uk/phs/mixed.

This book has been written to provide the reader with a thorough understanding of the concepts of mixed models and we trust it will serve well for this purpose. However, readers wishing to take a shortcut to the fitting of normal mixed models should read Chapter 1 for an introduction, Section 2.4 for practical details, and the chapter relevant to their design. To fit non-normal or categorical mixed models, Section 3.3 or Section 4.4 should be read in addition to Section 2.4. In an attempt to make this book easier to use, we have presented at the beginning of the text a summary of the notation we have used, while at the end we list some key definitions in a glossary.

Our writing of this book has been aided in many ways. The first edition evolved from a constantly changing set of course notes that accompanied a three-day course on the subject, run regularly over the previous six years. The second edition has been helped by many individuals who were kind enough to comment on the first edition, including the identification of some errors that had slipped in, and by further participants at our courses who have contributed to discussions and have thereby helped to shape our views. We are also grateful to many other colleagues who have read and commented on various sections of the manuscript and to our colleagues who have allowed us to use their data. We hope that readers will find the resulting book a useful reference in an interesting and expanding area of statistics.

Helen Brown
Robin Prescott
Edinburgh

Mixed Models Notation

The notation below is provided for quick reference. Models are defined more fully in Sections 2.1, 3.1 and 4.1.

Normal mixed model

$$
\begin{aligned}
&\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta} + \mathbf{e},\\
&\boldsymbol{\beta} \sim \mathrm{N}(\mathbf{0}, \mathbf{G}),\\
&\mathrm{var}(\mathbf{e}) = \mathbf{R},\\
&\mathrm{var}(\mathbf{y}) = \mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}.
\end{aligned}
$$

Generalised linear mixed model

$$
\begin{aligned}
&\mathbf{y} = \boldsymbol{\mu} + \mathbf{e},\\
&g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta},\\
&\boldsymbol{\beta} \sim \mathrm{N}(\mathbf{0}, \mathbf{G}),\\
&\mathrm{var}(\mathbf{e}) = \mathbf{R},\\
&\mathrm{var}(\mathbf{y}) = \mathbf{V} = \mathrm{var}(\boldsymbol{\mu}) + \mathbf{R},\\
&\qquad\qquad \approx \mathbf{BZGZ'B} + \mathbf{R} \quad \text{(a first-order approximation)},
\end{aligned}
$$

where

$\mathbf{y}$ = dependent variable,
$\mathbf{e}$ = residual error,
$\mathbf{X}$ = design matrix for fixed effects,
$\mathbf{Z}$ = design matrix for random effects,
$\boldsymbol{\alpha}$ = fixed effects parameters,
$\boldsymbol{\beta}$ = random effects parameters,
$\mathbf{R}$ = residual variance matrix,
$\mathbf{G}$ = matrix of covariance parameters,
$\mathbf{V} = \mathrm{var}(\mathbf{y})$ variance matrix,

$\boldsymbol{\mu}$ = expected values,
g = link function,
$\mathbf{B}$ = diagonal matrix of variance terms (e.g. $\mathbf{B} = \text{diag}\{\mu_i(1-\mu_i)\}$ for binary data).

Ordered categorical mixed model

$\mathbf{y} = \boldsymbol{\mu} + \mathbf{e}$,
$\text{logit}(\boldsymbol{\mu}^{[c]}) = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta}$,
$\boldsymbol{\beta} \sim \text{N}(\mathbf{0}, \mathbf{G})$,
var($\mathbf{y}$) is defined as in the GLMM,

where

$\boldsymbol{\mu} = (\mu_{11}, \mu_{12}, \mu_{13}, \mu_{21}, \mu_{22}, \mu_{23}, \ldots, \mu_{n1}, \mu_{n2}, \mu_{n3})'$,
μ_{ij} = probability observation i is in category j,
$\boldsymbol{\mu}^{[c]} = (\mu_{11}^{[c]}, \mu_{12}^{[c]}, \mu_{13}^{[c]}, \mu_{21}^{[c]}, \mu_{22}^{[c]}, \mu_{23}^{[c]}, \ldots, \mu_{n1}^{[c]}, \mu_{n2}^{[c]}, \mu_{n3}^{[c]})'$,
$\mu_{ij}^{[c]}$ = probability $(y_i \leq j) = \sum_{k=1}^{j} \mu_{ik}$.

Statistics in Practice

Human and Biological Sciences

Berger – Selection Bias and Covariate Imbalances in Randomized Clinical Trials
Brown and Prescott – Applied Mixed Models in Medicine, Second Edition
Chevret (Ed) – Statistical Methods for Dose-Finding Experiments
Ellenberg, Fleming and DeMets – Data Monitoring Committees in Clinical Trials: A Practical Perspective
Lawson, Browne and Vidal Rodeiro – Disease Mapping with WinBUGS and MLwiN
Lui – Statistical Estimation of Epidemiological Risk
*Marubini and Valsecchi – Analysing Survival Data from Clinical Trials and Observation Studies
Parmigiani – Modeling in Medical Decision Making: A Bayesian Approach
Senn – Cross-over Trials in Clinical Research, Second Edition
Senn – Statistical Issues in Drug Development
Spiegelhalter, Abrams and Myles – Bayesian Approaches to Clinical Trials and Health-Care Evaluation
Whitehead – Design and Analysis of Sequential Clinical Trials, Revised Second Edition
Whitehead – Meta-Analysis of Controlled Clinical Trials

Earth and Environmental Sciences

Buck, Cavanagh and Litton – Bayesian Approach to Interpreting Archaeological Data
Glasbey and Horgan – Image Analysis for the Biological Sciences
Helsel – Nondetects and Data Analysis: Statistics for Censored Environmental Data
McBride – Using Statistical Methods for Water Quality Management
Webster and Oliver – Geostatistics for Environmental Scientists

---

* Now available in paperback

Industry, Commerce and Finance

Aitken – Statistics and the Evaluation of Evidence for Forensic Scientists, Second Edition
Balding – Weight-of-evidence for Forensic DNA Profiles
Lehtonen and Pahkinen – Practical Methods for Design and Analysis of Complex Surveys, Second Edition
Ohser and Mücklich – Statistical Analysis of Microstructures in Materials Science
Taroni, Aitken, Garbolino and Biedermann – Bayesian Networks and Probabilistic Inference in Forensic Science

Contents